PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Should total calcium be adjusted for albumin? – a retrospective observational study of laboratory data from central Norway
AUTHORS	Lian, Ingrid; Åsberg, Arne

VERSION 1 – REVIEW

REVIEWER	Cristina Vassalle Fondazione CNR_Regione Toscana G Monasterio
REVIEW RETURNED	07-Jun-2017

GENERAL COMMENTS	The topic treated in the present manuscript is very interesting, and data clearly reported. - Better describe the population enrolled. How much healthy subjects? Which kind of patients?? Which diseases?? Critically ill patients were included? Were patients with primary hyperparathyroidism included??Add information on age, sex. -Any differences between inpatients and outpatients concerning the efficacy of parameters utilized? -Authors correctly used free calcium as gold standard. This fact imply that free calcium always reflects the “real” calcium levels, and that total calcium is incorrect when it disagrees with free calcium, which is reasonable but not necessarily always true. Moreover, free calcium may present limitations, due to improper technical issues, especially sample collection and handling. These points can be discussed in the discussion section.

REVIEWER	Javier Mateu - de Antonio Pharmacy Dept. Hospital del Mar Barcelona Spain
REVIEW RETURNED	09-Jun-2017

GENERAL COMMENTS	bmjopen-2017-017703 Total calcium should not be adjusted for albumin Review This is a retrospective study on a large laboratory data set from inpatients, outpatients and general practice patients in a hospital area in Norway. The objective was to test whether the local adjusted-calcium equation based on multiple linear regression improved compared to previously published formulas taking free calcium as a “gold standard” for calcium status.

	This article reinforces the idea of total calcium adjusted equations have low diagnostic accuracy. In my opinion, some points have to be explained to improve the paper: 1 – To my knowledge, the most accurate equation for corrected calcium and classify patients by calcium status is that published by James et al. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clinical Pathology 2008, 8:12 doi:10.1186/1472-6890-8-12. This equation was derived in the largest cohort up to date and validated in a large cohort as well. This equation should be included in the set of adjustment equations tested. 2 – As serum calcium levels could be related to severity of illness (see i.e.: Guven et al. Acta Neurol Belg. 2011 Mar;111(1):45-9 or Carlstedt et al. Eu J Clin Invest, 28: 898–903), it would be advisable to check the equation performance between inpatients vs. out-/ambulatory care patients (additionally in critically ill patients vs the rest of patients). It is possible that the accuracy of the equations varies significantly. 3 – What was the missing values treatment? 4 – Did authors used nonlinear equations in albumin < 27 g/L? Did they explored this approach to obtain general equations? To my knowledge, this approach (nonlinear) has been hardly used. 5 – In my view, some additional limitations of the study are the lack of some variables that could be significant for serum calcium (i.e. sodium, Mg or PTH), no comorbidities were recorded, relatively high proportion of creatinine alteration, and relatively young cohort.

REVIEWER	Catherine Clase McMaster University, Canada
REVIEW RETURNED	09-Jun-2017

GENERAL COMMENTS	Page numbers refer to the pdf numbering. **Major** This is an important topic, and a large dataset well suited to addressing it. The adjustment of calcium for albumin is a common clinical practice and accounts for an unknown, but likely important, proportion of albumin tests ordered, particularly repeated tests in hospitalized patients. The accuracy of the adjustment has been repeatedly questioned since 1978 but clinical practice, textbooks and guidelines have not changed. Accumulating data show that no adjustment formula improves on total calcium alone, and that the measurement properties of total calcium are fair to good. This work is the largest and best study that I am aware of in this area and could be practice changing. 1. Page 8 line 3. ``However, what the clinician really wants to know is how much the total concentration of calcium is`expected to change for one unit change in albumin concentration when the patient's condition is otherwise unchanged, specifically when the concentration of free calcium is unchanged.` et seq. This is clearly an important

point to the authors, but after several readings I do not understand it. What the clinician wants to know is the ionized calcium, and how it compares with the normal range, because that is what is biologically active. Attempts to calculate corrected calcium are just ways of trying to map the predicted ionized calcium onto a total calcium reference range. I suggest rewriting this paragraph.

2. Page 9 line 15. "We included data from both hospitalised patients and patients from outpatient clinics and general practice" Specimen handling is critical for ionized calcium results; samples are often drawn separately in blood gas syringes and should be sent on ice to the lab for immediate analysis. Please could you include more detail on how specimens were handled; if you have general policies around this it would be great to include them here, along with any QA information you have on how well they are followed. It would also be important to reflect critically on sample handling, especially from outside clinics, and it will likely be worthwhile to conduct a sensitivity analysis to examine any possible effect, and to mention handling as a possible limitation in the discussion.
3. The approach to accounting for renal insufficiency, dividing by creatinine above or below the upper limit of the lab normal range for creatinine (not given), is rather unsophisticated compared with current methods of classifying kidney function. I would suggest calculating eGFR by CKD-Epi.[1] (This requires knowledge of ethnic origin, usually not available in lab databases, but since it seems that around 1% of people in Sør-Trøndelag is of African origin (https://en.wikipedia.org/wiki/African_immigration_to_Norway), for the purposes of most epidemiological analyses, I think it would be reasonable to assess all participants as non-black.) If you wish to pre-specify cut points, I suggest using one of those suggested by KDIGO[2], perhaps 60 or 30 mL/min/1.73m^2 for this work.
4. Also specify whether creatinine is calibrated to isotope-dilution mass spectrometry. (I expect that it is, but if it is not, that does not invalidate the usefulness of the paper, given that the results are that no formula is an improvement on total; it would however reduce the generalizability of any formula derived here to other labs.)
5. page 10 line 4 et seq. The use of a different regression coefficients above and below an inflection point is fine, in modelling this is generally called a spline. (A nice example of this, showing the notation, can be found in the derivation of the CKD-Epi formula.[1]) You chose to use a single knot and the site of that based on graphical inspection, which I think is fine, though this can be done more objectively using advanced statistical methods. More problematic, if your formula were found to be useful, is the use of the age- and sex-specific creatinine cutpoints based on the lab's normal

ranges to determine where to place the knot(s) for creatinine. This is not the best approach to stratifying by renal function and would greatly complicate further application of the formula. I doubt that it affects your main finding - that even data-derived adjusting formulae doesn't improve on ionized calcium, but others might view this as a methodological flaw that would decrease the impact of your work. Consider redoing the analysis using spline methodology for albumin and for renal function; and use eGFR by CKD-Epi rather than creatinine as your measure of renal function.

6. I am not a statistician, but as I understand it, Harrell's C is a rank order concordance statistic developed for use in survival analysis. I'm not sure it is the best measure. When dealing with clinical lab results, it isn't just important to know whether one value is higher than another (does the test get the rank correct?) but the absolute value matters (what is the calcium in respect to the normal range?). To capture this, in our own work we have used the intraclass correlation coefficient, which assesses closeness to the line of identity. This requires converting both measures (ionized calcium and predicted total calcium) to a Z score because they are not inherently on the same scale.
7. The area under the curve statistics, however, are absolutely fine. The results are compelling and I doubt would be changed by attention to the details above: no prediction formula improves on the total calcium in predicting ionized hypo- or hyper-calcaemia.
8. Reporting the sensitivity, specificity, and perhaps likelihood ratios, for the detection of hypo- or hyper-calcaemia, by each of the methods would be of clinical utility.
9. Overall, I do think we need a really good paper on this question (previous work has not resulted in changes in practice or textbooks), and I think this could be it, but that would require attention to all the major details and some statistical reanalysis.
10. Conclusions, and conclusions of the abstract – that the common practice of adjusting calcium for albumin be abandoned – are important and justified.

Minor

11. The authors use the term 'free' rather than 'ionized' throughout. Ionized calcium would be my preferred term as this is how the lab test is named in English, at least wherever I have practised.
12. Include some of the actual numerical results for ICC and AUC in the abstract, since this will be what most people see.
13. Page 7 line 41. "a certain formula is only valid for specific patient populations [10]" Reference 10 is my paper: we certainly didn't argue that any formula was valid in any population. We found that albumin-corrected calcium performed poorly by all formulae in patients on dialysis and

	our review of the literature led us to note that no formula had been validated in any population as an improvement on total calcium. 14. Minor typographical errors and problems with agreement (eg, the word 'data' is plural). 15. Page 10, lines 27 to 54, some repetition. 16. Page 12. The BMJ formula is the one passed on from trainee to trainee and that appears in many textbooks. I'm very grateful to the authors for finding this original reference, I didn't know it before. I would include the formula in the paper at this point so that people understand the relevance of this analysis and why this is the graph you've chosen to include. 17. page 13, line 33 "Lastly, the reference limits of total calcium were better suited for unadjusted than for albumin-adjusted calcium (Figure 1)." I didn't understand this point and it doesn't seem to flow from the figure. 18. page 14, line 30. "In a position paper from 2006, the Kidney Disease: Improving Global Outcomes (KDIGO) acknowledged that calcium status is best monitored by measuring free calcium, but they also stated that if total calcium was used instead, it should be adjusted for low concentrations of albumin 22" The latest iteration is 2009 and the word 'possibly' has been introduced. http://kdigo.org/wp-content/uploads/2017/02/KDIGO-2009-CKD-MBD-Guideline-English.pdf 19. Thank you for the opportunity to review this interesting work, and hoping that it generates more interest than our similar recent paper![3] Our findings are entirely congruent with each other. Catherine Clase Associate professor, Nephrology, McMaster University Associate member, Health Research Methods, Evidence and Impact, McMaster University Deputy editor, Canadian Journal of Kidney Health and Disease 1. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. AnnInternMed 2009;150:604-12. 2. KDIGO CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3:1-150. 3. Steen OC, C.;Don-Wauchope A. Corrected Calcium Formula in Routine Clinical Use Does Not Accurately Reflect Ionized Calcium in Hospital Patients. Canadian Journal of General Internal Medicine 2015;10:14-21.

REVIEWER	Lars Mørkrid Department of Medical Biochemistry Oslo University Hospital NORWAY
REVIEW RETURNED	16-Jun-2017

GENERAL COMMENTS	My "NO"s in the list above are explained more in detail in the following text (also as a pdf.file enclosed). ********** General The authors have utilized retrospective laboratory production data after appropriate filtering (only the first sample from a patient is selected) to examine the relationship between P-free calcium as measured by a blood gas analyzer and total P-calcium and P-albumin. By multiple regression analysis with total calcium as the dependent variable and free calcium, albumin, phosphate, creatinine, sex and age as the explanatory variables they have obtained a “purified” albumin coefficient Calb, to establish the formula: Adjusted calcium = calcium + Calb × (40 – albumin). In the first place this has been done with data along the whole P-calcium range to compare it with formulas from other published studies. Later they performed the same type of calculations in four (somewhat arbitrarily chosen) subgroups, limited by different regions of covariate scales: albumin below and above 27 g/L in each subset of creatinine values below and above upper reference limit. The paper is well written and addresses a very relevant clinical issue, however the bombastic claim in the title needs to be somewhat modified until a more robust fundament for the gold standard (the reference interval of free calcium standardized at pH = 7.40) can be established. A more extensive discussion of what is new in this study as compared to other publications is also required. Special comments Page 6 line 35 1) The free calcium value standardized at pH = 7.40 is used in the calculations. Could that significantly affect the relationship between the other variables that have got their pattern of homeostatic balance at the actual pH? Page 6 line 42 2) The reference interval of free calcium lacks documentation. As the reference interval of free calcium is used as the gold standard, one might wonder if not the localization of the reference limits would greatly influence the course of and area under the ROC-curves, the values of the Harrell’s C index, as well as the curves representing central tendencies in Figure 1. 3) Has any important age dependency on free calcium been overlooked? Page 7 line 17 4) The choice of the fixed value 40 g/L in the equation has to be discussed.

	Page 7 line 24 5) As far as it can be inferred from the text, the values of coefficient Calb is obtained from linear multiple regression analysis, with backward elimination. Which p-value was used for the exclusion? Were the results compared with those obtained by another regression procedure, e.g. an enter method? 6) A curious reader may wonder if the same independent variables were thrown out in regressions for all four subgroups, and how big impact each of those retained might exert. Page 15 Referring to Table 1 the subdivisions need to be justified: 7) How can it be tested if the subdivisions into groups really result in regression coefficients that are statistically different? 8) If so, is the difference of biological or clinical importance? Page 16 The same type of argument as in point 8) above also applies to Table 2, but here the differences for the Harrell's C between "normal" and high creatinine values appear more "separated", as the mean estimate in the former group lies outside the 95% confidence interval in the latter and vice versa. 9) However this is not so easily seen when comparing line 1 in Table 2 (no adjustment) and line 2 (local adjustment). The authors have to explain if this is in agreement with the main conclusion of their paper. 10) The term normal creatinine is not ideal, as that subset also may contain pathological low values of creatinine, e.g. due to a low muscle mass etc.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer's comments on manuscript ID bmjopen-2017-017703 entitled "Total calcium should not be adjusted for albumin".

Editorial requests:

1) Please revise your title to indicate the research question, study design, and setting. This is the preferred format of the journal.

2) Thank you for providing the STARD checklist with your submission, however, you should have provided the STROBE checklist for the reporting of observational studies. Please in your next submission provide a completed STROBE checklist.

3) Please ensure that you discuss the strengths and limitations of your study design in the Discussion section.

Response to editorial requests:
1) The title has now been revised to "Should total calcium be adjusted for albumin? – a retrospective observational study of laboratory data from central Norway"

2) A complete STROBE checklist has been provided.

3) We have now included a more elaborate further discussion of the strengths and limitations of our study design in the Discussion section of the revised manuscript.

Response to Reviewer(s)' Comments:

Reviewer: 1
Cristina Vassalle

The topic treated in the present manuscript is very interesting, and data clearly reported.
1) Better describe the population enrolled. How much healthy subjects? Which kind of patients?? Which diseases?? Critically ill patients were included? Were patients with primary hyperparathyroidism included??Add information on age, sex.

Reply:
The population of 6567 patients included in this work was selected on the basis that they had measured total calcium, free calcium, creatinine, albumin and phosphate in the same blood draw. Our hospital is a full service acute care hospital that also analyses blood samples from general practice clinics in central Norway, and the population thus included samples from patients that were hospitalised, who visited the outpatient clinics, and patients from general practice where blood samples were sent to our lab for analysis.

As the data came from a laboratory database, we did not have access to diagnostic information. Only clinical data such as gender, sex and age was available to us. This information was described in the Material and Results section of the original manuscript. We have now extended the description of the population in the respective sections of the revised manuscript, and the available clinical information has been summarised in a new table, table 1 of the revised manuscript.

2) Any differences between inpatients and outpatients concerning the efficacy of parameters utilized?

Reply: To comply with other Reviewers' comments, we redid our analysis of the dataset, and included hospitalisation (or not) as dependent variables in the multiple linear regression analyses. As can be seen from the revised table 2, hospitalisation or not was only a significant variable in the subgroup of patients with eGFR ≥ 60 mL/minute/1.73 m2 and albumin ≥ 30 g/L (n=4910).

The population of inpatients in our study included only a very few critically ill patients, as free calcium in those patients were monitored using blood gas instruments in the intensive care units and the analytical results were not transferred to the laboratory information system.

We divided the population according to renal function (eGFR below or over 60) in the revised manuscript. A further division of those subgroups according to hospitalisation status did not appreciably change the results and conclusions. In the subgroup with eGFR 60, the formula of James et al. (1) was just a little better than total calcium in the hospitalised patients, but not in the ambulatory

patients. In the subgroup with eGFR < 60, total calcium was just a little better than the formula of James in the ambulant patients, but not in the hospitalised patients.

1) James MT, Zhang J, Lyon AW, Hemmelgarn BR. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clin Pathol. 2008;8:12.

3) Authors correctly used free calcium as gold standard. This fact imply that free calcium always reflects the "real" calcium levels, and that total calcium is incorrect when it disagrees with free calcium, which is reasonable but not necessarily always true. Moreover, free calcium may present limitations, due to improper technical issues, especially sample collection and handling. These points can be discussed in the discussion section.

Reply:
A discussion of the validity of pH-adjusted free calcium as the gold standard has now been included in the Discussion of the revised manuscript. In addition, a more detailed description of sample handling for analysis of free calcium analysis has been included in the Material and methods section of the revised manuscript, as this was requested by Reviewer #3. Of note, is that almost all samples consisted of venous blood drawn anaerobically into serum gel, such that sample collection was quite uniform across the population.

Reviewer: 2
Javier Mateu - de Antonio

This is a retrospective study on a large laboratory data set from inpatients, outpatients and general practice patients in a hospital area in Norway. The objective was to test whether the local adjusted-calcium equation based on multiple linear regression improved compared to previously published formulas taking free calcium as a "gold standard" for calcium status.
This article reinforces the idea of total calcium adjusted equations have low diagnostic accuracy. In my opinion, some points have to be explained to improve the paper:

1) To my knowledge, the most accurate equation for corrected calcium and classify patients by calcium status is that published by James et al. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clinical Pathology 2008, 8:12 doi:10.1186/1472-6890-8-12. This equation was derived in the largest cohort up to date and validated in a large cohort as well. This equation should be included in the set of adjustment equations tested.

Reply: The equation from James et al. has now been included in the set of adjustment equations tested in the revised manuscript.

2) As serum calcium levels could be related to severity of illness (see i.e.: Guven et al. Acta Neurol Belg. 2011 Mar;111(1):45-9 or Carlstedt et al. Eu J Clin Invest, 28: 898–903), it would be advisable to check the equation performance between inpatients vs. out-/ambulatory care patients (additionally in critically ill patients vs the rest of patients). It is possible that the accuracy of the equations varies significantly.

Reply: To comply with other Reviewers' comments, we redid our analysis of the dataset, and included hospitalisation (or not) as dependent variables in the multiple linear regression analyses. As can be seen from the revised table 2, hospitalisation or not was only a significant variable in the subgroup of patients with eGFR ≥ 60 mL/minute/1.73 m2 and albumin ≥ 30 g/L (n=4910).

The population of inpatients in our study included only a very few critically ill patients, as free calcium in those patients were monitored using blood gas instruments in the intensive care units and the analytical results were not transferred to the laboratory information system.

We divided the population according to renal function (eGFR below or over 60) in the revised manuscript. A further division of those subgroups according to hospitalisation status did not appreciably change the results and conclusions. In the subgroup with eGFR 60, the formula of James et al. (1) was just a little better than total calcium in the hospitalised patients, but not in the ambulatory patients. In the subgroup with eGFR < 60, total calcium was just a little better than the formula of James in the ambulant patients, but not in the hospitalised patients.

1) James MT, Zhang J, Lyon AW, Hemmelgarn BR. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clin Pathol. 2008;8:12.

3) What was the missing values treatment?

Reply: We only extracted laboratory data from patients where all the variables total calcium, free calcium, creatinine, albumin and phosphate (that had been analysed in a single blood draw) were present, and thus, no missing data treatment was necessary.

4) Did authors used nonlinear equations in albumin < 27 g/L? Did they explored this approach to obtain general equations? To my knowledge, this approach (nonlinear) has been hardly used.

Reply: First of all, our dataset has been reanalysed due to suggestions from several Reviewers. The linear regression analysis has now been performed in subgroups with albumin below or over 30 g/L in the revised manuscript.
More specifically, the total dataset was divided in three subgroups. First, according to eGFR below or above 60 mL/minute/1.73 m2, as others have found different albumin coefficients in individuals with renal failure compared to individuals with normal renal function (1). Then, for patients with eGFR ≥ 60, we divided the dataset according to albumin concentrations below or above 30 g/L, as locally weighted scatterplot smoothing of total calcium against albumin indicated nonlinearity overall, but linearity below and above 30 g/L (see figures below). We also formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin 30 g/L. For patients with eGFR > 60, the slope was statistically significantly larger for the subgroup with albumin < 30 than for the subgroup with albumin 30 ($p < 0.001$). No such difference was found for patients with eGFR < 60 ($p = 0.934$).

An albumin coefficient was then calculated for each of the three subgroups, and used in the evaluation of the diagnostic accuracy of our local formula. This has now been described more carefully in the revised manuscript.

eGFR < 60 eGFR ≥ 60

(1) Ceriotti F, Boyd JC, Klein G, Henny J, Queralto J, Kairisto V, et al. Reference intervals for serum creatinine concentrations: assessment of available data for global application. Clin Chem. 2008;54:559-66.

5) In my view, some additional limitations of the study are the lack of some variables that could be significant for serum calcium (i.e. sodium, Mg or PTH), no comorbidities were recorded, relatively high proportion of creatinine alteration, and relatively young cohort.

Reply: We agree that additional laboratory variables and clinical information would have been useful. However, the inclusion of such variables would significantly have reduced the size of the dataset. We wanted to keep the samples size as large as possible to get a reliable estimate of the albumin coefficient. The relatively large number of patients with reduced renal function in the study population may be an advantage, as we know from this and other studies that albumin-adjustment formulas perform differently in patients with and without renal failure. These limitations have now been included in the Discussion of the revised manuscript.

Reviewer: 3
Catherine Clase
Page numbers refer to the pdf numbering.

Major
This is an important topic, and a large dataset well suited to addressing it. The adjustment of calcium for albumin is a common clinical practice and accounts for an unknown, but likely important, proportion of albumin tests ordered, particularly repeated tests in hospitalized patients. The accuracy of the adjustment has been repeatedly questioned since 1978 but clinical practice, textbooks and guidelines have not changed. Accumulating data show that no adjustment formula improves on total calcium alone, and that the measurement properties of total calcium are fair to good. This work is the largest and best study that I am aware of in this area and could be practice changing.

1. Page 8 line 3. ``However, what the clinician really wants to know is how much the total concentration of calcium is`expected to change for one unit change in albumin concentration when the patient's condition is otherwise unchanged, specifically when the concentration of free calcium is unchanged.` et seq. This is clearly an important point to the authors, but after several readings I do not understand it. What the clinician wants to know is the ionized calcium, and how it compares with the normal range, because that is what is biologically active. Attempts to calculate corrected calcium are just ways of trying to map the predicted ionized calcium onto a total calcium reference range. I suggest rewriting this paragraph.
Reply: We agree that this point is somewhat unclear, and it has been changed in the revised version of the manuscript.

In performing albumin adjustments of calcium, one is indeed trying to find a “surrogate” for free calcium. We believe that the traditional approach of regressing the concentration of total calcium against albumin using simple linear regression is wrong. Instead, one should use multiple linear regression to find the expected change in calcium for one unit change in albumin concentration per se, holding the other variables such as albumin, free calcium, creatinine, phosphate sex, age etc. constant; to find albumins ”eigen-effect”, ie. more ”information” that albumin alone is necessary in such regression analyses., as these variables affect the relationship between albumin and calcium.

The paragraph has now been changed to:
"However, when making an albumin-adjustment we should use a coefficient that shows how much the total concentration of calcium is expected to change for one unit change in albumin concentration when the patient's condition is otherwise unchanged, specifically when the concentration of free calcium is unchanged. To estimate that coefficient we have to regress the concentration of total calcium against albumin and free calcium, sex, age or whatever explanatory variable is relevant. Then the interpretation of the albumin coefficient gets in line with its use."

2. Page 9 line 15. "We included data from both hospitalised patients and patients from outpatient clinics and general practice" Specimen handling is critical for ionized calcium results; samples are often drawn separately in blood gas syringes and should be sent on ice to the lab for immediate analysis. Please could you include more detail on how specimens were handled; if you have general policies around this it would be great to include them here, along with any QA information you have on how well they are followed. It would also be important to reflect critically on sample handling, especially from outside clinics, and it will likely be worthwhile to conduct a sensitivity analysis to examine any possible effect, and to mention handling as a possible limitation in the discussion.
Reply: Additional details on sample handling for the analysis of free calcium at our laboratory has been included in the Material and methods section of the revised manuscript. In addition, the validity of pH-adjusted free calcium as the gold standard has now been discussed in the Discussion of the revised manuscript. Unfortunately, no QA information on routine follow-up was available. Of note, almost all samples from our population consisted of venous blood drawn anaerobically into serum gel. Thus, sample collection was quite uniform across our population.

3. The approach to accounting for renal insufficiency, dividing by creatinine above or below the upper limit of the lab normal range for creatinine (not given), is rather unsophisticated compared with current methods of classifying kidney function. I would suggest calculating eGFR by CKD-Epi.1 (This requires knowledge of ethnic origin, usually not available in lab databases, but since it seems that around 1% of people in Sør-Trøndelag is of African origin (https://en.wikipedia.org/wiki/African_immigration_to_Norway), for the purposes of most epidemiological analyses, I think it would be reasonable to assess all participants as non-black.) If you wish to pre-specify cut points, I suggest using one of those suggested by KDIGO2, perhaps 60 or 30 mL/min/1.73m2 for this work.
Reply:
We agree with the Reviewer. We did not use eGFR in the original manuscript, as there has been no available formula for individuals <18 years based solely on age, gender and creatinine concentration, to our knowledge. However, we became aware of a recent publication (1) with an eGFR formula, the full age spectrum (FAS) equation, that was validated for both children (above 2 years of age) and adults. This formula was used to account for renal insufficiency (we divided our dataset into subgroups with eGFR below or above 60 mL/minute/1.73 m2) in the revised manuscript.

(1) Pottel H, Hoste L, Dubourg L, Ebert N, Schaeffner E, Eriksen BO, et al. An estimated glomerular filtration rate equation for the full age spectrum. Nephrol Dial Transplant. 2016;31:798-806.

4. Also specify whether creatinine is calibrated to isotope-dilution mass spectrometry. (I expect that it is, but if it is not, that does not invalidate the usefulness of the paper, given that the results are

that no formula is an improvement on total; it would however reduce the generalizability of any formula derived here to other labs.)
Reply: The creatinine assay was an enzymatic method calibrated against an isotope dilution mass spectrometry (IDMS) reference method. This information has now been included in the revised manuscript.

5. page 10 line 4 et seq. The use of a different regression coefficients above and below an inflection point is fine, in modelling this is generally called a spline. (A nice example of this, showing the notation, can be found in the derivation of the CKD-Epi formula.1) You chose to use a single knot and the site of that based on graphical inspection, which I think is fine, though this can be done more objectively using advanced statistical methods. More problematic, if your formula were found to be useful, is the use of the age- and sex-specific creatinine cut points based on the lab's normal ranges to determine where to place the knot(s) for creatinine. This is not the best approach to stratifying by renal function and would greatly complicate further application of the formula. I doubt that it affects your main finding - that even data-derived adjusting formulae doesn't improve on ionized calcium, but others might view this as a methodological flaw that would decrease the impact of your work. Consider redoing the analysis using spline methodology for albumin and for renal function; and use eGFR by CKD-Epi rather than creatinine as your measure of renal function.
Reply: As mentioned above, the dataset has now been reanalysed, subgrouped according to eGFR rather than creatinine concentrations, as suggested. We visually checked the lowess lines in plot of total calcium against albumin separately for patients with eGFR ≥ 60 and eGFR < 60, and fond a break-point at albumin of 30 g/L in the group with eGFR ≥ 60, but not in the other group. Then we formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin ≥ 30 g/L. For patients with eGFR>60, the slope was statistically significantly larger for the subgroup with albumin<30 than for the subgroup with albumin ≥ 30 ($p < 0.001$). No such difference was found for patients with eGFR<60 ($p = 0.934$).

6. I am not a statistician, but as I understand it, Harrell's C is a rank order concordance statistic developed for use in survival analysis. I'm not sure it is the best measure. When dealing with clinical lab results, it isn't just important to know whether one value is higher than another (does the test get the rank correct?) but the absolute value matters (what is the calcium in respect to the normal range?). To capture this, in our own work we have used the intraclass correlation coefficient, which assesses closeness to the line of identity. This requires converting both measures (ionized calcium and predicted total calcium) to a Z score because they are not inherently on the same scale.
Reply: In this work, we used Harrell's c index as a measure of diagnostic accuracy. This index is related to the area under the ROC curve. Both measures are 0.5 at no diagnostic accuracy and 1.0 at perfect diagnostic accuracy. Harrell's c takes on the same value as the area under the ROC curve when the gold standard is binary.

We believe Harrell`s c index is a useful measure of diagnostic accuracy, because it "improves" on the ROC analysis as it contains more information, by using free calcium as a continuous gold standard, as opposed to a dichotomous gold standard in ROC curve analysis. We have included ROC curve analysis as well, as it is may be more familiar to readers in evaluating diagnostic accuracy. Furthermore, we have expanded the ROC curve analyses to account for different definitions/limits of hypo- and hypercalcemia in the revised manuscript. This has been clarified in the revised manuscript.

7. The area under the curve statistics, however, are absolutely fine. The results are compelling and I doubt would be changed by attention to the details above: no prediction formula improves on the total calcium in predicting ionized hypo- or hyper-calcaemia.
Reply: We agree.

8. Reporting the sensitivity, specificity, and perhaps likelihood ratios, for the detection of hypo- or hyper-calcaemia, by each of the methods would be of clinical utility.
Reply: We disagree slightly. We have focused on the overall diagnostic accuracy, as given by Harrell´s c and area under ROC curve. We are not sure that clinicians use data on sensitivity and specificity of albumin-adjusted calcium or unadjusted calcium.

9. Overall, I do think we need a really good paper on this question (previous work has not resulted in changes in practice or textbooks), and I think this could be it, but that would require attention to all the major details and some statistical reanalysis.
Reply: We agree with Reviewer #3, and hope that this revised paper might convince clinical doctors to to change their current practice.

10. Conclusions, and conclusions of the abstract – that the common practice of adjusting calcium for albumin be abandoned – are important and justified.
Reply: We thank Reviewer #3 for this comment.

Minor
11. The authors use the term 'free' rather than 'ionized' throughout. Ionized calcium would be my preferred term as this is how the lab test is named in English, at least wherever I have practised.
Reply: We do not agree with Reviewer #3 on this. All calcium atoms in the body are ionized, but in plasma, only 50% of the calcium ions are free/unbound and ready to exert biological effects, and the rest are bound to proteins and in complexes. We therefore believe the term "free calcium" is the most correct term for unbound calcium ions, despite that the term "ionized calcium" is now commonly used.

12. Include some of the actual numerical results for ICC and AUC in the abstract, since this will be what most people see.
Reply: We have now included numerical results for diagnostic accuracy (Harrell's c) in the abstract of the revised manuscript.

13. Page 7 line 41. "a certain formula is only valid for specific patient populations 10" Reference 10 is my paper: we certainly didn't argue that any formula was valid in any population. We found that albumin-corrected calcium performed poorly by all formulae in patients on dialysis and our review of the literature led us to note that no formula had been validated in any population as an improvement on total calcium.

Reply: We apologize for our error in referring to your work in this sentence, it is an error on our part, and we have now corrected the reference.

14. Minor typographical errors and problems with agreement (eg, the word 'data' is plural).
Reply: We thank Reviewer #3 for the observation, corrections have been made accordingly.

15. Page 10, lines 27 to 54, some repetition.
Reply: This has now been rewritten in the revised version of the manuscript.

16. Page 12. The BMJ formula is the one passed on from trainee to trainee and that appears in many textbooks. I'm very grateful to the authors for finding this original reference, I didn't know it before. I would include the formula in the paper at this point so that people understand the relevance of this analysis and why this is the graph you've chosen to include.
Reply: The formula has now been included in the revised manuscript. We agree with the Reviewer that this could be helpful for the reader.

17. page 13, line 33 "Lastly, the reference limits of total calcium were better suited for unadjusted than for albumin-adjusted calcium (Figure 1)." I didn't understand this point and it doesn't seem to flow from the figure.
Reply: We agree with Reviewer #3 that this sentence is somewhat unclear. After a reanalysis of our dataset, we chose to exclude this figure.

18. page 14, line 30. "In a position paper from 2006, the Kidney Disease: Improving Global Outcomes (KDIGO) acknowledged that calcium status is best monitored by measuring free calcium, but they also stated that if total calcium was used instead, it should be adjusted for low concentrations of albumin 22" The latest iteration is 2009 and the word 'possibly' has been introduced. http://kdigo.org/wp-content/uploads/2017/02/KDIGO-2009-CKD-MBD-Guideline-English.pdf
Reply: We thank Reviewer #3 for the updated information. After a revision of the manuscript, this paragraph and the attached references has been deleted.

19. Thank you for the opportunity to review this interesting work, and hoping that it generates more interest than our similar recent paper!3 Our findings are entirely congruent with each other.
Reply: We thank the Reviewer for the work performed, and the corrections and insightful comments and suggestions. We share the hope that this work will be reach a broad audience, and possibly change current practice.

Catherine Clase
Associate professor, Nephrology, McMaster University
Associate member, Health Research Methods, Evidence and Impact, McMaster University

Deputy editor, Canadian Journal of Kidney Health and Disease

1. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. AnnInternMed 2009;150:604-12.
2. KDIGO CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3:1-150.
3. Steen OC, C.;Don-Wauchope A. Corrected Calcium Formula in Routine Clinical Use Does Not Accurately Reflect Ionized Calcium in Hospital Patients. Canadian Journal of General Internal Medicine 2015;10:14-21.

Reviewer: 4
Lars Mørkrid

General:
The authors have utilized retrospective laboratory production data after appropriate filtering (only the first sample from a patient is selected) to examine the relationship between P-free calcium as measured by a blood gas analyzer and total P-calcium and P-albumin. By multiple regression analysis with total calcium as the dependent variable and free calcium, albumin, phosphate, creatinine, sex and age as the explanatory variables they have obtained a “purified” albumin coefficient Calb, to establish the formula: Adjusted calcium = calcium + Calb × (40 – albumin). In the first place this has been done with data along the whole P-calcium range to compare it with formulas from other published studies. Later they performed the same type of calculations in four (somewhat arbitrarily chosen) subgroups, limited by different regions of covariate scales: albumin below and above 27 g/L in each subset of creatinine values below and above upper reference limit.

The paper is well written and addresses a very relevant clinical issue, however the bombastic claim in the title needs to be somewhat modified until a more robust fundament for the gold standard (the reference interval of free calcium standardized at pH = 7.40) can be established.

Reply: The title of the manuscript has now been changed to ”Should total calcium be adjusted for albumin – a retrospective observational study of laboratory data from central Norway”.

A more extensive discussion of what is new in this study as compared to other publications is also required.

Reply: We have now included a more extensive discussion of what is new in this study as compared to previous publications in the Discussion of the revised manuscript.

Special comments
Page 6 line 35
1) The free calcium value standardized at pH = 7.40 is used in the calculations. Could that significantly affect the relationship between the other variables that have got their pattern of homeostatic balance at the actual pH?

Reply: As far as we know, the other explanatory variables (age, creatinine/eGFR, albumin and phosphate) are not pH-dependent, so we do not quite understand the question.

Of interest, we have now included a discussion of the validity of pH-adjusted free calcium as the gold standard in the revised manuscript.

Page 6 line 42
2) The reference interval of free calcium lacks documentation. As the reference interval of free calcium is used as the gold standard, one might wonder if not the localization of the reference limits would greatly influence the course of and area under the ROC-curves, the values of the Harrell's C index, as well as the curves representing central tendencies in Figure 1.

Reply: We thank the Reviewer for this comment. We have now included documentation of the reference interval in the revised manuscript. Furthermore, we have expanded the ROC curve analyses to account for different definitions/limits of hypo- and hypercalcemia, see figure 1 of the revised manuscript (below). The reference limits of hypo- and hypercalcemia does however not influence the value of the Harrell´s c index.

3) Has any important age dependency on free calcium been overlooked?

Reply: When we introduced an interaction term between free calcium and age in the multiple regression analyses, the final albumin coefficient did not change for patients with eGFR < 60 and for patients with eGFR > 60 and albumin < 30. For patients with eGFR > 60 and albumin > 30, the albumin coefficient increased from 0.1204 to 0.1207 when the interaction term was included; however, we did not think this was relevant. In general, we did not explore the possible effect of the numerous interaction terms between the explanatory variables.

Page 7 line 17
4) The choice of the fixed value 40 g/L in the equation has to be discussed.

Reply: The various adjustment-formulas use different normal values of albumin. We normalised to 40 g/L. The choice of normal albumin value does not influence the diagnostic accuracy, because adjusted calcium = calcium + coefficient × (normal albumin - albumin) = calcium + coefficient × normal albumin - coefficient × albumin = calcium + constant + coefficient × albumin. Adding a constant to the value of a diagnostic marker does not change its diagnostic accuracy. The choice of normal albumin value does, however, influence the optimal cut-off value of albumin-adjusted calcium. Finding the optimal cut-off value could be done by ROC analysis if the prevalence of the clinical condition and the consequences of false and true positive and negative results are known, but such an analysis was beyond the scope of this work. This has now been discussed in the revised manuscript.

Page 7 line 24
5) As far as it can be inferred from the text, the values of coefficient Calb is obtained from linear multiple regression analysis, with backward elimination. Which p-value was used for the exclusion? Were the results compared with those obtained by another regression procedure, e.g. an enter method?

Reply: A $p < 0.05$ was chosen for inclusion of variables in the multiple linear regression. This information has now been included in the revised manuscript under Material and methods, statistical analysis. The results were not compared to other regression procedures.

6) A curious reader may wonder if the same independent variables were thrown out in regressions for all four subgroups, and how big impact each of those retained might exert.

Reply: In the revised manuscript, we reanalysed the dataset due to suggestions from several Reviewers. We now have three subgroups in the regression analyses. In the multiple regression analyses, we included free calcium, albumin, phosphate, eGFR, gender, age and hospitalisation (or not) as the explanatory/independent variables. It varied between subgroups which variables were retained. The retained variables are listed in table 2 of the revised manuscript. In addition, the results of univariate analysis are also presented, so the reader can see the effect of including other variables than just albumin. In our opinion, a more detailed presentation is not feasible.

Page 15 Referring to Table 1 the subdivisions need to be justified:
7) How can it be tested if the subdivisions into groups really result in regression coefficients that are statistically different?

Reply: As we have reanalysed the dataset, a more detailed and hopefully clear description of the statistical analyses, including the justification for the subgrouping of the population, has been provided. Specifically, in the simple linear regression analysis of calcium against albumin, we formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin 30 g/L. For patients with eGFR > 60, the slope was statistically significantly larger for the subgroup with albumin < 30 g/L than for the subgroup with albumin 30 g/L ($p < 0.001$). No such difference was found for patients with eGFR < 60 ($p = 0.934$).

8) If so, is the difference of biological or clinical importance?

Reply: We are not sure what the Reviewer #4 means by this. The justification for dividing the population/dataset into subgroups has been included in the revised manuscript, hopefully this is clarifying. How to evaluate the clinical importance of a certain difference in albumin coefficient is not clear to us. Anyway, such a task was beyond the scope of this work.

Page 16
The same type of argument as in point 8) above also applies to Table 2, but here the differences for the Harrell's C between "normal" and high creatinine values appear more "separated", as the mean estimate in the former group lies outside the 95% confidence interval in the latter and vice versa.

Reply: A new analysis of the dataset has been performed, and the Harrell´s c estimates were tested against each other using the" lincom" procedure. The results, with relevant p - values, have now been reported in the revised manuscript.

9) However this is not so easily seen when comparing line 1 in Table 2 (no adjustment) and line 2 (local adjustment). The authors have to explain if this is in agreement with the main conclusion of their paper.

Reply: As mentioned above, a new analysis of the dataset has been performed, and the Harrell´s c estimates were tested against each. The results, with relevant p - values, have now been reported in the revised manuscript. Hopefully, the results in the revised manuscript are clearer.

10) The term normal creatinine is not ideal, as that subset also may contain pathological low values of creatinine, e.g. due to a low muscle mass etc.

Reply: The term ”normal creatinine” has now been removed from the revised manuscript. We have now have reanalysed the dataset using eGFR according to a formula that is validated both children (above 2 years of age) and adults. This formula was used to account for renal insufficiency (we divided our dataset into subgroups with eGFR below or above 60 mL/minute/1.73 m2) in the revised manuscript.

(1) Pottel H, Hoste L, Dubourg L, Ebert N, Schaeffner E, Eriksen BO, et al. An estimated glomerular filtration rate equation for the full age spectrum. Nephrol Dial Transplant. 2016;31:798-806.

VERSION 2 – REVIEW

REVIEWER	Javier Mateu-de Antonio, Pharm D Hospital del Mar Barcelona Spain
REVIEW RETURNED	19-Sep-2017

GENERAL COMMENTS	Review Authors have made an extensive revision, including statistics, of their paper attending reviewer's comments. In my opinion, the work has improved as the most of queries has been correctly addressed. Limitations are clearly detailed. Graphics are neat and tables easy understandable. My recommendation is that the paper could be published.

REVIEWER	Lars Mørkrid Department of Medical Biochemistry, Oslo University Hospital, Oslo, Norway
REVIEW RETURNED	03-Oct-2017

GENERAL COMMENTS	The authors have responded satisfactorily to most of the objections raised in my primary review. However, concerning objection #2, the traceability of reference interval of free calcium still needs some clarification. Referring to page 50 lines 27-28, reference 13 is not easily accessible, and therefore a short description of the reference population (recruitment, inclusion/exclusion criteria, number) and the statistical

	method is essential. Referring to objection #8 one should not automatically consider a statistically significant difference as biologically or clinically significant. To which extent an "effect size" of a subdivision is large enough to be practically relevant is important in order to simplify models. E.g. in establishing reference limits between genders Gowans' criterion or partition rules may be used for that purpose. For the present publication I do not demand the authors to perform such analyses (this may be a rather complex task with their data), but they should at least mention it to keep caution about such phenomena.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer's comments on manuscript ID bmjopen-2017-017703 entitled *"Total calcium should not be adjusted for albumin".*

Editorial requests:

1) Please revise your title to indicate the research question, study design, and setting. This is the preferred format of the journal.

2) Thank you for providing the STARD checklist with your submission, however, you should have provided the STROBE checklist for the reporting of observational studies. Please in your next submission provide a completed STROBE checklist.

3) Please ensure that you discuss the strengths and limitations of your study design in the Discussion section.

Response to editorial requests:

1) The title has now been revised to "*Should total calcium be adjusted for albumin? – a retrospective observational study of laboratory data from central Norway*"

2) A complete STROBE checklist has been provided.

3) We have now included a more elaborate further discussion of the strengths and limitations of our study design in the Discussion section of the revised manuscript.

Response to Reviewer(s)' Comments:

Reviewer: 1

Cristina Vassalle

The topic treated in the present manuscript is very interesting, and data clearly reported.

1) Better describe the population enrolled. How much healthy subjects? Which kind of patients?? Which diseases?? Critically ill patients were included? Were patients with primary

hyperparathyroidism included??Add information on age, sex.

Reply:

The population of 6567 patients included in this work was selected on the basis that they had measured total calcium, free calcium, creatinine, albumin and phosphate in the same blood draw. Our hospital is a full service acute care hospital that also analyses blood samples from general practice clinics in central Norway, and the population thus included samples from patients that were hospitalised, who visited the outpatient clinics, and patients from general practice where blood samples were sent to our lab for analysis.

As the data came from a laboratory database, we did not have access to diagnostic information. Only clinical data such as gender, sex and age was available to us. This information was described in the Material and Results section of the original manuscript. We have now extended the description of the population in the respective sections of the revised manuscript, and the available clinical information has been summarised in a new table, table 1 of the revised manuscript.

2) Any differences between inpatients and outpatients concerning the efficacy of parameters utilized*?*

Reply: To comply with other Reviewers' comments, we redid our analysis of the dataset, and included hospitalisation (or not) as dependent variables in the multiple linear regression analyses. As can be seen from the revised table 2, hospitalisation or not was only a significant variable in the subgroup of patients with eGFR ≥ 60 mL/minute/1.73 m^2 and albumin ≥ 30 g/L (n=4910).

The population of inpatients in our study included only a very few critically ill patients, as free calcium in those patients were monitored using blood gas instruments in the intensive care units and the analytical results were not transferred to the laboratory information system.

We divided the population according to renal function (eGFR below or over 60) in the revised manuscript. A further division of those subgroups according to hospitalisation status did not appreciably change the results and conclusions. In the subgroup with eGFR ≥ 60, the formula of James et al. (1) was just a little better than total calcium in the hospitalised patients, but not in the ambulatory patients. In the subgroup with eGFR < 60, total calcium was just a little better than the formula of James in the ambulant patients, but not in the hospitalised patients.

1) James MT, Zhang J, Lyon AW, Hemmelgarn BR. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clin Pathol. 2008;8:12.

3) Authors correctly used free calcium as gold standard. This fact imply that free calcium always reflects the "real" calcium levels, and that total calcium is incorrect when it disagrees with free calcium, which is reasonable but not necessarily always true. Moreover, free calcium may present limitations, due to improper technical issues, especially sample collection and handling. These points can be discussed in the discussion section.

Reply:

A discussion of the validity of pH-adjusted free calcium as the gold standard has now been included in the Discussion of the revised manuscript. In addition, a more detailed description of sample handling

for analysis of free calcium analysis has been included in the Material and methods section of the revised manuscript, as this was requested by Reviewer #3. Of note, is that almost all samples consisted of venous blood drawn anaerobically into serum gel, such that sample collection was quite uniform across the population.

Reviewer: 2

Javier Mateu - de Antonio

This is a retrospective study on a large laboratory data set from inpatients, outpatients and general practice patients in a hospital area in Norway. The objective was to test whether the local adjusted-calcium equation based on multiple linear regression improved compared to previously published formulas taking free calcium as a "gold standard" for calcium status.

This article reinforces the idea of total calcium adjusted equations have low diagnostic accuracy. In my opinion, some points have to be explained to improve the paper:

1) To my knowledge, the most accurate equation for corrected calcium and classify patients by calcium status is that published by James et al. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clinical Pathology 2008, 8:12 doi:10.1186/1472-6890-8-12. This equation was derived in the largest cohort up to date and validated in a large cohort as well. This equation should be included in the set of adjustment equations tested.

Reply: The equation from James et al. has now been included in the set of adjustment equations tested in the revised manuscript.

2) As serum calcium levels could be related to severity of illness (see i.e.: Guven et al. Acta Neurol Belg. 2011 Mar;111(1):45-9 or Carlstedt et al. Eu J Clin Invest, 28: 898–903), it would be advisable to check the equation performance between inpatients vs. out-/ambulatory care patients (additionally in critically ill patients vs the rest of patients). It is possible that the accuracy of the equations varies significantly.

Reply: To comply with other Reviewers' comments, we redid our analysis of the dataset, and included hospitalisation (or not) as dependent variables in the multiple linear regression analyses. As can be seen from the revised table 2, hospitalisation or not was only a significant variable in the subgroup of patients with eGFR ≥ 60 mL/minute/1.73 m^2 and albumin ≥ 30 g/L (n=4910).

The population of inpatients in our study included only a very few critically ill patients, as free calcium in those patients were monitored using blood gas instruments in the intensive care units and the analytical results were not transferred to the laboratory information system.

We divided the population according to renal function (eGFR below or over 60) in the revised manuscript. A further division of those subgroups according to hospitalisation status did not appreciably change the results and conclusions. In the subgroup with eGFR ≥ 60, the formula of

James et al. (1) was just a little better than total calcium in the hospitalised patients, but not in the ambulatory patients. In the subgroup with eGFR < 60, total calcium was just a little better than the formula of James in the ambulant patients, but not in the hospitalised patients.

1) James MT, Zhang J, Lyon AW, Hemmelgarn BR. Derivation and internal validation of an equation for albumin-adjusted calcium. BMC Clin Pathol. 2008;8:12.

3) What was the missing values treatment?

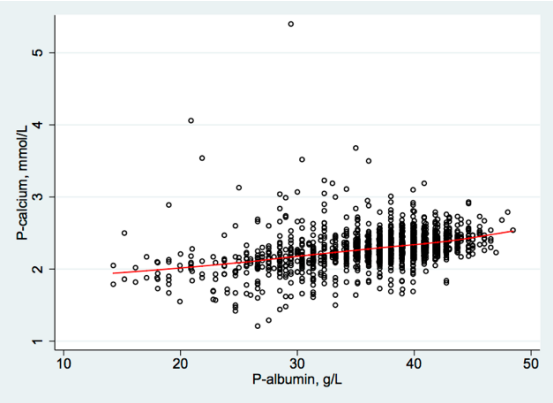
Reply: We only extracted laboratory data from patients where all the variables total calcium, free calcium, creatinine, albumin and phosphate (that had been analysed in a single blood draw) were present, and thus, no missing data treatment was necessary.

4) Did authors used nonlinear equations in albumin < 27 g/L? Did they explored this approach to obtain general equations? To my knowledge, this approach (nonlinear) has been hardly used.

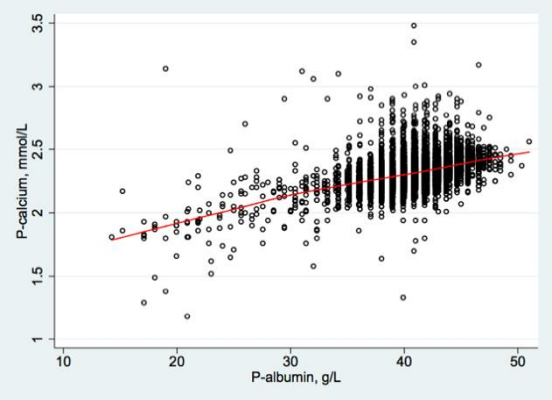
Reply: First of all, our dataset has been reanalysed due to suggestions from several Reviewers. The linear regression analysis has now been performed in subgroups with albumin below or over 30 g/L in the revised manuscript.

More specifically, the total dataset was divided in three subgroups. First, according to eGFR below or above 60 mL/minute/1.73 m^2, as others have found different albumin coefficients in individuals with renal failure compared to individuals with normal renal function (1). Then, for patients with eGFR ≥ 60, we divided the dataset according to albumin concentrations below or above 30 g/L, as locally weighted scatterplot smoothing of total calcium against albumin indicated nonlinearity overall, but linearity below and above 30 g/L (see figures below). We also formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin ≥ 30 g/L. For patients with eGFR > 60, the slope was statistically significantly larger for the subgroup with albumin < 30 than for the subgroup with albumin ≥ 30 ($p < 0.001$). No such difference was found for patients with eGFR < 60 ($p = 0.934$).

An albumin coefficient was then calculated for each of the three subgroups, and used in the evaluation of the diagnostic accuracy of our local formula. This has now been described more carefully in the revised manuscript.



eGFR < 60



eGFR ≥ 60

(1) Ceriotti F, Boyd JC, Klein G, Henny J, Queralto J, Kairisto V, et al. Reference intervals for serum creatinine concentrations: assessment of available data for global application. Clin Chem. 2008;54:559-66.

5) In my view, some additional limitations of the study are the lack of some variables that could be significant for serum calcium (i.e. sodium, Mg or PTH), no comorbidities were recorded, relatively high proportion of creatinine alteration, and relatively young cohort.

Reply: We agree that additional laboratory variables and clinical information would have been useful. However, the inclusion of such variables would significantly have reduced the size of the dataset. We wanted to keep the samples size as large as possible to get a reliable estimate of the albumin coefficient. The relatively large number of patients with reduced renal function in the study population may be an advantage, as we know from this and other studies that albumin-adjustment formulas perform differently in patients with and without renal failure. These limitations have now been included in the Discussion of the revised manuscript.

Reviewer: 3

Catherine Clase

Page numbers refer to the pdf numbering.

Major

This is an important topic, and a large dataset well suited to addressing it. The adjustment of calcium for albumin is a common clinical practice and accounts for an unknown, but likely important, proportion of albumin tests ordered, particularly repeated tests in hospitalized patients. The accuracy of the adjustment has been repeatedly questioned since 1978 but clinical practice, textbooks and guidelines have not changed. Accumulating data show that no adjustment formula improves on total calcium alone, and that the measurement properties of total calcium are fair to good**.** This work is the largest and best study that I am aware of in this area and could be practice changing.

20. Page 8 line 3. ``However, what the clinician really wants to know is how much the total concentration of calcium is`expected to change for one unit change in albumin concentration when the patient's condition is otherwise unchanged, specifically when the concentration of free calcium is unchanged.` et seq. This is clearly an important point to the authors, but after several readings I do not understand it. What the clinician wants to know is the ionized calcium, and how it compares with the normal range, because that is what is biologically

active. Attempts to calculate corrected calcium are just ways of trying to map the predicted ionized calcium onto a total calcium reference range. I suggest rewriting this paragraph.

Reply: We agree that this point is somewhat unclear, and it has been changed in the revised version of the manuscript.

In performing albumin adjustments of calcium, one is indeed trying to find a "surrogate" for free calcium. We believe that the traditional approach of regressing the concentration of total calcium against albumin using simple linear regression is wrong. Instead, one should use multiple linear regression to find the expected change in calcium for one unit change in albumin concentration *per se*, holding the other variables such as albumin, free calcium, creatinine, phosphate sex, age etc. constant; to find albumins "eigen-effect", ie. more "information" that albumin alone is necessary in such regression analyses., as these variables affect the relationship between albumin and calcium.

The paragraph has now been changed to:

"However, when making an albumin-adjustment we should use a coefficient that shows how much the total concentration of calcium is expected to change for one unit change in albumin concentration when the patient's condition is otherwise unchanged, specifically when the concentration of free calcium is unchanged. To estimate that coefficient we have to regress the concentration of total calcium against albumin and free calcium, sex, age or whatever explanatory variable is relevant. Then the interpretation of the albumin coefficient gets in line with its use."

21. Page 9 line 15. "We included data from both hospitalised patients and patients from outpatient clinics and general practice" Specimen handling is critical for ionized calcium results; samples are often drawn separately in blood gas syringes and should be sent on ice to the lab for immediate analysis. Please could you include more detail on how specimens were handled; if you have general policies around this it would be great to include them here, along with any QA information you have on how well they are followed. It would also be important to reflect critically on sample handling, especially from outside clinics, and it will likely be worthwhile to conduct a sensitivity analysis to examine any possible effect, and to mention handling as a possible limitation in the discussion.

Reply: Additional details on sample handling for the analysis of free calcium at our laboratory has been included in the Material and methods section of the revised manuscript. In addition, the validity of pH-adjusted free calcium as the gold standard has now been discussed in the Discussion of the revised manuscript. Unfortunately, no QA information on routine follow-up was available. Of note, almost all samples from our population consisted of venous blood drawn anaerobically into serum gel. Thus, sample collection was quite uniform across our population.

22. The approach to accounting for renal insufficiency, dividing by creatinine above or below the upper limit of the lab normal range for creatinine (not given), is rather unsophisticated compared with current methods of classifying kidney function. I would suggest calculating eGFR by CKD-Epi.[1] (This requires knowledge of ethnic origin, usually not available in lab databases, but since it seems that around 1% of people in Sør-Trøndelag is of African origin (https://en.wikipedia.org/wiki/African_immigration_to_Norway), for the purposes of most epidemiological analyses, I think it would be reasonable to assess all participants as non-black.) If you wish to pre-specify cut points, I suggest using one of those suggested by KDIGO[2], perhaps 60 or 30 mL/min/1.73m^2 for this work.

Reply:

We agree with the Reviewer. We did not use eGFR in the original manuscript, as there has been no available formula for individuals <18 years based solely on age, gender and creatinine concentration, to our knowledge. However, we became aware of a recent publication (1) with an eGFR formula, the full age spectrum (FAS) equation, that was validated for both children (above 2 years of age) and adults. This formula was used to account for renal insufficiency (we divided our dataset into subgroups with eGFR below or above 60 mL/minute/1.73 $m^{2)}$ in the revised manuscript.

(1) Pottel H, Hoste L, Dubourg L, Ebert N, Schaeffner E, Eriksen BO, et al. An estimated glomerular filtration rate equation for the full age spectrum. Nephrol Dial Transplant. 2016;31:798-806.

23. Also specify whether creatinine is calibrated to isotope-dilution mass spectrometry. (I expect that it is, but if it is not, that does not invalidate the usefulness of the paper, given that the results are that no formula is an improvement on total; it would however reduce the generalizability of any formula derived here to other labs.)

Reply: The creatinine assay was an enzymatic method calibrated against an isotope dilution mass spectrometry (IDMS) reference method. This information has now been included in the revised manuscript.

24. page 10 line 4 et seq. The use of a different regression coefficients above and below an inflection point is fine, in modelling this is generally called a spline. (A nice example of this, showing the notation, can be found in the derivation of the CKD-Epi formula.[1]) You chose to use a single knot and the site of that based on graphical inspection, which I think is fine, though this can be done more objectively using advanced statistical methods. More problematic, if your formula were found to be useful, is the use of the age- and sex-specific creatinine cut points based on the lab's normal ranges to determine where to place the knot(s) for creatinine. This is not the best approach to stratifying by renal function and would greatly complicate further application of the formula. I doubt that it affects your main finding - that even data-derived adjusting formulae doesn't improve on ionized calcium, but others might view this as a methodological flaw that would decrease the impact of your work. Consider redoing the analysis using spline methodology for albumin and for renal function; and use eGFR by CKD-Epi rather than creatinine as your measure of renal function.

Reply: As mentioned above, the dataset has now been reanalysed, subgrouped according to eGFR rather than creatinine concentrations, as suggested. We visually checked the lowess lines in plot of total calcium against albumin separately for patients with eGFR ≥ 60 and eGFR < 60, and fond a break-point at albumin of 30 g/L in the group with eGFR ≥ 60, but not in the other group. Then we formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin ≥ 30 g/L. For patients with eGFR>60, the slope was statistically significantly larger for the subgroup with albumin<30 than for the subgroup with albumin ≥ 30 ($p < 0.001$). No such difference was found for patients with eGFR<60 ($p = 0.934$).

25. I am not a statistician, but as I understand it, Harrell's C is a rank order concordance statistic developed for use in survival analysis. I'm not sure it is the best measure. When dealing with clinical lab results, it isn't just important to know whether one value is higher than another (does the test get the rank correct?) but the absolute value matters (what is the calcium in respect to the normal range?). To capture this, in our own work we have used the intraclass correlation coefficient, which assesses closeness to the line of identity. This requires converting both measures (ionized calcium and predicted total calcium) to a Z score because they are not inherently on the same scale.

Reply: In this work, we used Harrell's c index as a measure of diagnostic accuracy. This index is related to the area under the ROC curve. Both measures are 0.5 at no diagnostic accuracy and 1.0 at perfect diagnostic accuracy. Harrell's c takes on the same value as the area under the ROC curve when the gold standard is binary.

We believe Harrell`s c index is a useful measure of diagnostic accuracy, because it "improves" on the ROC analysis as it contains more information, by using free calcium as a *continuous* gold standard, as opposed to a *dichotomous* gold standard in ROC curve analysis. We have included ROC curve analysis as well, as it is may be more familiar to readers in evaluating diagnostic accuracy. Furthermore, we have expanded the ROC curve analyses to account for different definitions/limits of hypo- and hypercalcemia in the revised manuscript. This has been clarified in the revised manuscript.

26. The area under the curve statistics, however, are absolutely fine. The results are compelling and I doubt would be changed by attention to the details above: no prediction formula improves on the total calcium in predicting ionized hypo- or hyper-calcaemia.

Reply: We agree.

27. Reporting the sensitivity, specificity, and perhaps likelihood ratios, for the detection of hypo- or hyper-calcaemia, by each of the methods would be of clinical utility.

Reply: We disagree slightly. We have focused on the overall diagnostic accuracy, as given by Harrell´s c and area under ROC curve. We are not sure that clinicians use data on sensitivity and specificity of albumin-adjusted calcium or unadjusted calcium.

28. Overall, I do think we need a really good paper on this question (previous work has not resulted in changes in practice or textbooks), and I think this could be it, but that would require attention to all the major details and some statistical reanalysis.

Reply: We agree with Reviewer #3, and hope that this revised paper might convince clinical doctors to to change their current practice.

29. Conclusions, and conclusions of the abstract – that the common practice of adjusting calcium for albumin be abandoned – are important and justified.

Reply: We thank Reviewer #3 for this comment.

Minor

30. The authors use the term 'free' rather than 'ionized' throughout. Ionized calcium would be my preferred term as this is how the lab test is named in English, at least wherever I have practised.

Reply: We do not agree with Reviewer #3 on this. All calcium atoms in the body are ionized, but in plasma, only 50% of the calcium ions are free/unbound and ready to exert biological effects, and the rest are bound to proteins and in complexes. We therefore believe the term "free calcium" is the most correct term for unbound calcium ions, despite that the term "ionized calcium" is now commonly used.

31. Include some of the actual numerical results for ICC and AUC in the abstract, since this will be what most people see.

Reply: We have now included numerical results for diagnostic accuracy (Harrell's c) in the abstract of the revised manuscript.

32. Page 7 line 41. "a certain formula is only valid for specific patient populations [10]" Reference 10 is my paper: we certainly didn't argue that any formula was valid in any population. We found that albumin-corrected calcium performed poorly by all formulae in patients on dialysis

and our review of the literature led us to note that no formula had been validated in any population as an improvement on total calcium.

Reply: We apologize for our error in referring to your work in this sentence, it is an error on our part, and we have now corrected the reference.

33. Minor typographical errors and problems with agreement (eg, the word 'data' is plural).

Reply: We thank Reviewer #3 for the observation, corrections have been made accordingly.

34. Page 10, lines 27 to 54, some repetition.

Reply: This has now been rewritten in the revised version of the manuscript.

35. Page 12. The BMJ formula is the one passed on from trainee to trainee and that appears in many textbooks. I'm very grateful to the authors for finding this original reference, I didn't know it before. I would include the formula in the paper at this point so that people understand the relevance of this analysis and why this is the graph you've chosen to include.

Reply: The formula has now been included in the revised manuscript. We agree with the Reviewer that this could be helpful for the reader.

36. page 13, line 33 "Lastly, the reference limits of total calcium were better suited for unadjusted than for albumin-adjusted calcium (Figure 1)." I didn't understand this point and it doesn't seem to flow from the figure.

Reply: We agree with Reviewer #3 that this sentence is somewhat unclear. After a reanalysis of our dataset, we chose to exclude this figure.

37. page 14, line 30. "In a position paper from 2006, the Kidney Disease: Improving Global Outcomes (KDIGO) acknowledged that calcium status is best monitored by measuring free calcium, but they also stated that if total calcium was used instead, it should be adjusted for low concentrations of albumin 22" The latest iteration is 2009 and the word 'possibly' has been introduced. http://kdigo.org/wp-content/uploads/2017/02/KDIGO-2009-CKD-MBD-Guideline-English.pdf

Reply: We thank Reviewer #3 for the updated information. After a revision of the manuscript, this paragraph and the attached references has been deleted.

38. Thank you for the opportunity to review this interesting work, and hoping that it generates more interest than our similar recent paper![3] Our findings are entirely congruent with each other.

Reply: We thank the Reviewer for the work performed, and the corrections and insightful comments and suggestions. We share the hope that this work will be reach a broad audience, and possibly change current practice.

Catherine Clase

Associate professor, Nephrology, McMaster University

Associate member, Health Research Methods, Evidence and Impact, McMaster University

Deputy editor, Canadian Journal of Kidney Health and Disease

1. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. AnnInternMed 2009;150:604-12.

2. KDIGO CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3:1-150.

3. Steen OC, C.;Don-Wauchope A. Corrected Calcium Formula in Routine Clinical Use Does Not Accurately Reflect Ionized Calcium in Hospital Patients. Canadian Journal of General Internal Medicine 2015;10:14-21.

Reviewer: 4

Lars Mørkrid

General:

The authors have utilized retrospective laboratory production data after appropriate filtering

(only the first sample from a patient is selected) to examine the relationship between P-free calcium as measured by a blood gas analyzer and total P-calcium and P-albumin. By multiple regression analysis with total calcium as the dependent variable and free calcium, albumin, phosphate, creatinine, sex and age as the explanatory variables they have obtained a "purified" albumin coefficient Calb, to establish the formula: Adjusted calcium = calcium + Calb × (40 – albumin). In the first place this has been done with data along the whole P-calcium range to compare it with formulas from other published studies. Later they performed the same type of calculations in four (somewhat arbitrarily chosen) subgroups, limited by different regions of covariate scales: albumin below and above 27 g/L in each subset of creatinine values below and above upper reference limit.

The paper is well written and addresses a very relevant clinical issue, however the bombastic claim in the title needs to be somewhat modified until a more robust fundament for the gold standard (the reference interval of free calcium standardized at pH = 7.40) can be established.

Reply: The title of the manuscript has now been changed to *"Should total calcium be adjusted for albumin – a retrospective observational study of laboratory data from central Norway".*

A more extensive discussion of what is new in this study as compared to other publications is also required.

Reply: We have now included a more extensive discussion of what is new in this study as compared to previous publications in the Discussion of the revised manuscript.

Special comments

Page 6 line 35

1) The free calcium value standardized at pH = 7.40 is used in the calculations. Could that significantly affect the relationship between the other variables that have got their pattern of homeostatic balance at the actual pH?

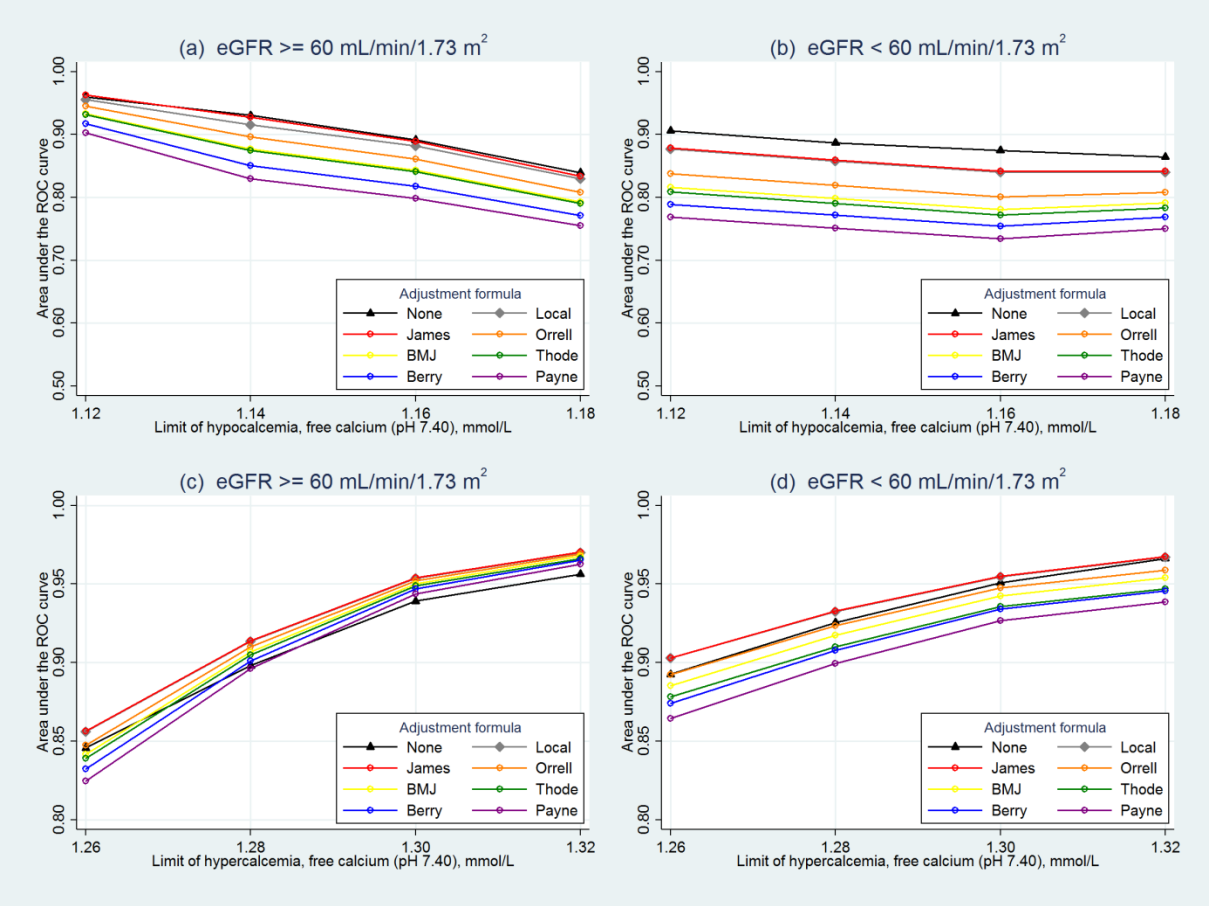
Reply: As far as we know, the other explanatory variables (age, creatinine/eGFR, albumin and phosphate) are not pH-dependent, so we do not quite understand the question.

Of interest, we have now included a discussion of the validity of pH-adjusted free calcium as the gold standard in the revised manuscript.

Page 6 line 42

2) The reference interval of free calcium lacks documentation. As the reference interval of free calcium is used as the gold standard, one might wonder if not the localization of the reference limits would greatly influence the course of and area under the ROC-curves, the values of the Harrell's C index, as well as the curves representing central tendencies in Figure 1.

Reply: We thank the Reviewer for this comment. We have now included documentation of the reference interval in the revised manuscript. Furthermore, we have expanded the ROC curve analyses to account for different definitions/limits of hypo- and hypercalcemia, see figure 1 of the revised manuscript (below). The reference limits of hypo- and hypercalcemia does however not influence the value of the Harrell´s c index.



3) Has any important age dependency on free calcium been overlooked?

Reply: When we introduced an interaction term between free calcium and age in the multiple regression analyses, the final albumin coefficient did not change for patients with eGFR < 60 and for patients with eGFR > 60 and albumin < 30. For patients with eGFR > 60 and albumin > 30, the albumin coefficient increased from 0.1204 to 0.1207 when the interaction term was included; however, we did not think this was relevant. In general, we did not explore the possible effect of the

numerous interaction terms between the explanatory variables.

Page 7 line 17

4) The choice of the fixed value 40 g/L in the equation has to be discussed.

Reply: The various adjustment-formulas use different normal values of albumin. We normalised to 40 g/L. The choice of normal albumin value does not influence the diagnostic accuracy, because adjusted calcium = calcium + coefficient × (normal albumin - albumin) = calcium + coefficient × normal albumin - coefficient × albumin = calcium + constant + coefficient × albumin. Adding a constant to the value of a diagnostic marker does not change its diagnostic accuracy. The choice of normal albumin value does, however, influence the optimal cut-off value of albumin-adjusted calcium. Finding the optimal cut-off value could be done by ROC analysis if the prevalence of the clinical condition and the consequences of false and true positive and negative results are known, but such an analysis was beyond the scope of this work. This has now been discussed in the revised manuscript.

Page 7 line 24

5) As far as it can be inferred from the text, the values of coefficient Calb is obtained from linear multiple regression analysis, with backward elimination. Which p-value was used for the exclusion? Were the results compared with those obtained by another regression procedure, e.g. an enter method?

Reply: A $p < 0.05$ was chosen for inclusion of variables in the multiple linear regression. This information has now been included in the revised manuscript under Material and methods, statistical analysis. The results were not compared to other regression procedures.

6) A curious reader may wonder if the same independent variables were thrown out in regressions for all four subgroups, and how big impact each of those retained might exert.

Reply: In the revised manuscript, we reanalysed the dataset due to suggestions from several Reviewers. We now have three subgroups in the regression analyses. In the multiple regression analyses, we included free calcium, albumin, phosphate, eGFR, gender, age and hospitalisation (or not) as the explanatory/independent variables. It varied between subgroups which variables were retained. The retained variables are listed in table 2 of the revised manuscript. In addition, the results of univariate analysis are also presented, so the reader can see the effect of including other variables than just albumin. In our opinion, a more detailed presentation is not feasible.

Page 15 Referring to Table 1 the subdivisions need to be justified:

7) How can it be tested if the subdivisions into groups really result in regression coefficients that are statistically different?

Reply: As we have reanalysed the dataset, a more detailed and hopefully clear description of the statistical analyses, including the justification for the subgrouping of the population, has been provided. Specifically, in the simple linear regression analysis of calcium against albumin, we formally tested whether the slope of the regression line was different for albumin < 30 g/L compared to albumin ≥ 30 g/L. For patients with eGFR > 60, the slope was statistically significantly larger for the subgroup with albumin < 30 g/L than for the subgroup with albumin ≥ 30 g/L ($p < 0.001$). No such difference was found for patients with eGFR < 60 ($p = 0.934$).

8) If so, is the difference of biological or clinical importance?

Reply: We are not sure what the Reviewer #4 means by this. The justification for dividing the population/dataset into subgroups has been included in the revised manuscript, hopefully this is

clarifying. How to evaluate the clinical importance of a certain difference in albumin coefficient is not clear to us. Anyway, such a task was beyond the scope of this work.

Page 16

The same type of argument as in point 8) above also applies to Table 2, but here the differences for the Harrell's C between "normal" and high creatinine values appear more "separated", as the mean estimate in the former group lies outside the 95% confidence interval in the latter and vice versa.

Reply: A new analysis of the dataset has been performed, and the Harrell´s c estimates were tested against each other using the" lincom" procedure. The results, with relevant p - values, have now been reported in the revised manuscript.

9) However this is not so easily seen when comparing line 1 in Table 2 (no adjustment) and line 2 (local adjustment). The authors have to explain if this is in agreement with the main conclusion of their paper.

Reply: As mentioned above, a new analysis of the dataset has been performed, and the Harrell´s c estimates were tested against each. The results, with relevant p - values, have now been reported in the revised manuscript. Hopefully, the results in the revised manuscript are clearer.

10) The term normal creatinine is not ideal, as that subset also may contain pathological low values of creatinine, e.g. due to a low muscle mass etc.

Reply: The term "normal creatinine" has now been removed from the revised manuscript. We have now have reanalysed the dataset using eGFR according to a formula that is validated both children (above 2 years of age) and adults. This formula was used to account for renal insufficiency (we divided our dataset into subgroups with eGFR below or above 60 mL/minute/1.73 m^2) in the revised manuscript.

(1) Pottel H, Hoste L, Dubourg L, Ebert N, Schaeffner E, Eriksen BO, et al. An estimated glomerular filtration rate equation for the full age spectrum. Nephrol Dial Transplant. 2016;31:798-806.

VERSION 3 – REVIEW

REVIEWER	Lars Mørkrid Department of Medical Biochemistry Rikshospitalet - Oslo University Hospital Oslo NORWAY
REVIEW RETURNED	11-Dec-2017

GENERAL COMMENTS	As noted before, the gold standard (the traceability of free calcium) still poses a problem. The authors have adequately responded to my earlier objection and provided an accesible link to Ref. # 13 (in Danish): Bruunhuus I, Magdal U. Kompendium 2000 - Kompendium i Laboratoriemedicin online [Edited 2008]. Available from: http://www.dskb.dk/Clubs/CommonDrive/Components/GetWWWFile.aspx?fileID=45556. However, as far I can see, this does not give any information about the reference population, instrument used for the measurements and statistical method for calculation of the reference interval, which could easily been stated by adding a couple of lines with this information to the section Reference ranges on Page 8. (Or is it something that I may have overlooked?). With this solved, points 8, 11 and 12 in the checklist above can be completed. _

	I would defer to more qualified language experts to judge checkpoint 15.

VERSION 3 – AUTHOR RESPONSE

Reply to reviwer #4:

Unfortunately, the online Danish reference #13 of the reference interval for ionized calcium does not contain any description of the reference population, the instrument used for the measurements or the statistical method for calculation of the reference interval. This is the same reference interval as referred to in “Kompendium i laboratoriemedicin, Ed. Henrik Olesen, Amtrådsforeningen, 1988.”, this information is not given here either. Rigshospitalet in Denmark uses the very same reference interval, (http://labportal.rh.dk/Metodeliste.asp?Mode=Display&Id=2311) without stating the source. Therefore, we are unfortunately not able to provide the requested information in the manuscript.

Nonetheless, this is the reference interval our lab uses, and as mentioned earlier; we have adressed the question of the localization of the reference limits in a previous Reply to the reviewers; where we expanded the ROC curve analyses to account for different definitions/limits of hypo- and hypercalcemia, see figure 1 of the previously revised manuscript. The reference limits of hypo- and hypercalcemia does not influence the value of the Harrell´s c index.